External Information in Community Detection

Weston D. Viles* A. James O'Malley†

Abstract
The literature on network science abounds with diverse community detection methods. Included among them, modularity optimization has found widespread application and success. From a statistical physics perspective, modularity optimization and its variants are equivalent to ground state determination of the Potts model. We extend this notion to include the external field in the Potts energy function to impose constraints on the derived communities. The incorporation of external information of a global nature has remained largely unstudied. We retain the probabilistic aspect of the Potts model with an external field to define an MCMC sampling regime for the community labels. We apply this novel method to a network of hospitals, some of which are specialized. Our constraint is that each *health care community* (HCC) must contain at least one specialized hospital. For one to make meaningful comparisons among discovered communities, a level of homogeneity of resource availability is necessary. The specialized hospitals are equipped to implant implantable cardiac defibrillators (ICDs) and, while all hospitals are partitioned into HCCs, we require that each HCC contain at least one ICD capable hospital.

1. Introduction

Network analysis methods are widely employed by researchers intent on gaining a systems-level understanding of relational and dependent data. Technological and infrastructure, social, biological, and information sciences are a few of the major disciplines in which network analyses have been successfully employed. Graph partitioning, or community detection[1], has experienced several landmark advances in methodology, chief among them the clique percolation method [5], spectral partitioning [3], degree-corrected stochastic block models [11], and modularity optimization [12]. These tools are specifically designed for the unsupervised partitioning of the vertex set V of a network graph into unusually cohesive subsets of vertices. With varied applications as sociology [20], computer architecture [8], and biology [14], graph partitioning procedures are fast becoming indispensable devices for scientific research.

Existing methods of community detection incorporate solely the network's topology in partitioning. In modularity optimization, for example, an objective function is defined on the network and is maximized over the latent partition labels in order to discover cohesive subsets of vertices. Unfortunately, direct application of a network modularity optimization procedure is not particularly amenable to the imposition of constraints or incorporation of external information on the communities themselves. In this article we propose a procedure for obtaining high-quality, feasible communities by integrating constraints on the partitions discovered by our method.

Our motivation is derived from our desire to partition a network of hospitals into *health care communities* (HCCs). Hospitals constituting an HCC should exhibit a high frequency of shared visits amongst one-another within the Medicare health service and are densely connected in the network graph representation of the overall hospital network. A proper

*Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, 878 Rubin, Dartmouth-Hitchcock Medical Center, 1 Medical Center Drive, Lebanon, NH 03756

†The Dartmouth Institute, Geisel School of Medicine, Dartmouth College, Room 544, Williamson Translational Research Building, 1 Medical Center Drive, Lebanon, NH 03756

[1] here we do not differentiate between graph partitioning and community detection, see [12] for details.

partitioning of the hospital network into HCCs requires that each HCC consist of at least one specialized cardiac hospital known as a *cardiac care facility*. This additional information is not intrinsic to the network connectivity itself and, therefore, must be directly incorporated into the community detection procedure. Existing work in this realm considers incorporating additional information of the forms of individual entity labels and pairwise constraints, i.e. that two vertices must be labelled similarly or differently, see [4]. Our additional information instead imposes a constraint on the composition of network communities and, therefore, requires a novel approach be developed.

We derive our constrained community detection method from the first principles of statistical physics to obtain a probabilistic model for the joint distribution of the latent community labels and an associated sampling method. For illustrative purposes, suppose that we aim to partition the vertex set into only two communities so that a latent community label is a binary random variable. Note here that this is the exact scenario in Zachary's Karate Club network. A general probabilistic model for the joint distribution of a collection of binary random variables is the *Ising model*, see [9]. Specifically, suppose that $x = (x_!, x_2, \ldots, x_p)$ is a *configuration*, i.e. instance or realization, of p binary variables. The Ising model is given by

$$\mathbb{P}\left(x\right) \propto \exp\left\{-\beta\left[\mu\sum_{k} h_k x_k + \sum_{ij} J_{ij} x_i x_j\right]\right\}, \tag{1}$$

where $\beta > 0$ and $\mu \in \mathbb{R}$ are constant, h_k is the *external magnetic field* applied to x_k, and J_{ij} is the *interaction strength* between x_i and x_j. The Ising model may be written succinctly as

$$\mathbb{P}\left(x\right) \propto \exp\left\{-\beta\mathcal{H}\left(x\right)\right\}, \tag{2}$$

where $\mathcal{H}\left(x\right)$ is the *energy function*, or *Hamiltonian*, of the Ising model. In general, Equation (2) is the form of a *Gibbs distribution*.

The energy function of our Gibbs distribution is composed, as above, of two parts: (1) the external field which incorporates outside knowledge and (2) the interaction component which assimilates the network connectivity, i.e. $\mathcal{H}\left(x\right) = \mu\mathcal{H}_E\left(x\right) + \mathcal{H}_I\left(x\right)$, where $\mathcal{H}_E\left(x\right)$ is the external field and $\mathcal{H}_I\left(x\right)$ is the interaction component. Proceeding with the Ising model illustration,

$$\mathcal{H}_E\left(x\right) = \sum_{k} h_k x_k \tag{3}$$

and

$$\mathcal{H}_I\left(x\right) = \sum_{ij} J_{ij} x_i x_j. \tag{4}$$

The *external field* is aptly named because it quantifies the role of forces external to the interactions within the system. We utilize the role of the external field as the courier of external information on the latent network communities. We demonstrate our community detection method using illustrative examples and subsequently employ our method in partitioning a nationwide hospital network.

The organization of this article is as follows. In Section 2 we begin by presenting the perhaps most commonly used approach to community detection, modularity for optimization, and conclude with the development of our constrained objective function and allied optimization procedure. In Section 3 we present a numerical experiment to demonstrate the functioning of the external field in our constrained probabilistic optimization problem. In Section 5 we apply our method to partition the hospital referral network into HCCs and compare the result to the existing partition known as *health referral regions* (HRRs), see [7]. Technical details and derivations appear in the Appendix.

2. Function Optimization and Probability Models for Community Detection

Modularity optimization, see [6] and [12], is a procedure for community detection in which the maximum of an objective function establishes the optimal partition of the vertex set V. The modularity objective function is defined specifically for the purpose of grouping together vertices which are better connected than expected and grouping seperately vertices which are lesser connected than expected. This, of course, presupposes the notion of expected connectivity and, thus, a null probability distribution over the network, see the Appendix for details.

Network modularity is exclusively a function of a network's *adjacency matrix*. An adjacency matrix, in general, stores edge weights. In the case that the network is *simple*, *unweighted*, and *undirected*, as are the networks we consider here, then the adjacency matrix is a symmetric binary matrix consisting of only zeros and ones. Specifically,

$$A_{ij} = \begin{cases} 1 & \text{if vertices } v_i \text{ and } v_j \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The definition of the modularity function Q, see [6] and [12], is

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \frac{s_i s_j + 1}{2}, \tag{6}$$

where $2m = \sum_{ij} A_{ij}$ is twice the number of edges in the network (m is the number of edges), d_k is the *degree* of vertex v_k, and

$$s_k = \begin{cases} 1 & \text{if } v_k \text{ belongs to Community \#1} \\ -1 & \text{if } v_k \text{ belongs to Community \#2} \end{cases} \tag{7}$$

The degree d_k of vertex v_k is equal to the number of edges connected to v_k and equals the k^{th} column (row) total, i.e. $d_k = \sum_j A_{kj}$. Equation (6) is easily modified to accommodate several communities by redefining Q as

$$Q = \frac{1}{4m} \sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \left[2\delta\left(x_i, x_j\right) - 1 \right], \tag{8}$$

where x_k is the community label for vertex v_k and $\delta\left(\cdot, \cdot\right)$ is the Kronecker delta.

The vector $x = (x_1, x_2, \ldots, x_p)$ of community labels, where p is the number of vertices in the network, is the variable over which Q is optimized to obtain the optimal community assignment. Note that the role of the scaling factor $1/4m$ does not enter into the optimization. In fact, neither do the terms with $i = j$. These terms are referred to as diagonal because they are along the main diagonal of the outer product $A - \mathbf{d}\mathbf{d}^T/2m$, where $\mathbf{d}^T$ is the transpose of $\mathbf{d} = (d_1, d_2, \ldots, d_p)^T$.

Thus, we define Q' as

$$Q' = \sum_{i \neq j} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \left[2\delta\left(x_i, x_j\right) - 1 \right] \tag{9}$$

as the quantity to be optimized. Fundamental to the remainder of this article is the observation that Equation (9) is the Hamiltonian of a Potts (Ising) model.

We refer the reader to the Appendix for two derivations of the modularity function.

2.1 The Probabilistic Model of Constrained Communities

Having identified the objective function Q' of a network partition, we seek an optimization procedure to compute optimal community labels given the network's connectivity. However, if information beyond the connectivity of the network is to be incorporated in assigning community labels, the objective function Q' for network modularity must be modified.

The authors of [4] proposed a semi-supervised method of constrained community detection that directly alters the null model of the modularity function. Specifically, they incorporate external knowledge of the form of (i) individual entity labels and (ii) pairwise constraints, i.e. that two vertices must be labeled similarly or differently. Instead of placing local constraints on a desired solution to the modularity problem, we search for a solution which satisfies a global constraint. Global constraint satisfaction may be achieved through a proper coding of the external field, which we present in the following. The incorporation of external information of this nature has, to our knowledge, remained unstudied.

As stated earlier, Q' can be recognized as the Hamiltonian of an Ising or Potts model so we shall now write

$$\mathcal{H}_I\left(x\right) = \sum_{i \neq j} \left(A_{ij} - \frac{d_i d_j}{2m}\right)\left[2\delta\left(x_i, x_j\right) - 1\right], \tag{10}$$

where the subscript I indicates "interaction". As discussed earlier, in the definition of $\mathcal{H}_I$ the diagonal terms in Q' are omitted since they amount to a constant not depending upon the community labels. The probabilistic model is thus

$$\mathbb{P}\left(x\right) \propto \exp\left\{\theta\left[\lambda \mathcal{H}_E\left(x\right) + \mathcal{H}_I\left(x\right)\right]\right\}, \tag{11}$$

for some penalty parameter $\lambda \geq 0$ and inverse temperature $\theta > 0$, where the function $\mathcal{H}_E$ is the external field of the system. As in Section 1, the model in Equation (11) may be written succinctly as

$$\mathbb{P}\left(x\right) \propto \left\{\theta \mathcal{H}\left(x\right)\right\}, \tag{12}$$

where the Hamiltonian $\mathcal{H} = \lambda \mathcal{H}_E\left(x\right) + \mathcal{H}_I\left(x\right)$. It is clear that $\lambda = 0$ implies that no penalty is imposed. The external field is characterized by knowledge in the form of constraints that are not directly encoded in the connectivity of the network. Moreover, external information is often in the form of a constraint on the composition of network communities.

Consider the following example: Suppose there exist p vertices in a network and that r of the p vertices are *special vertices*, vertices whose community labels are relevant to a constraint $\mathcal{C}$ positing that each discovered community contain at least one such special vertex. Clearly, the number of communities $|C|$ necessarily is such that $|C| \leq r$.

For illustration, define a binary function $\chi\left(x_k\right)$ of the label x_k of vertex v_k as $\chi\left(x_k\right) =$ 1 if

- $v_k \in SV$ and
- there exists a community with no special vertices and
- there exist ≥ 2 special vertices in the community to which v_k belongs,

and 0 otherwise. It is natural then to define

$$\mathcal{H}_E\left(x\right) = \sum_k e_k \chi\left(x_k\right) \tag{13}$$

where the e_k are chosen to balance the corresponding interaction terms in $\mathcal{H}_I$. To this end, note that

$$\begin{aligned}\mathcal{H}_I(x) &= 2\sum_{i=2}^{p-1}\sum_{j=i+1}^{p}\left(A_{ij}-\frac{d_i d_j}{2m}\right)\left[2\delta\left(x_i,x_j\right)-1\right] \\ &\quad +2\sum_{j=2}^{p}\left(A_{1j}-\frac{d_1 d_j}{2m}\right)\left[2\delta\left(x_1,x_2\right)-1\right] \qquad (14)\end{aligned}$$

so that the total contribution of all terms involving vertex v_1 is

$$2\sum_{j=2}^{p}\left(A_{1j}-\frac{d_1 d_j}{2m}\right)=2\frac{d_1^2}{2m}. \qquad (15)$$

The total contribution of vertex v_1 may be interpreted as the value when all vertices are labeled identically. It is interesting to note that, for the class of simple graphs which we consider, this is, up to the factor 2, the diagonal terms we discarded from Q' in defining $\mathcal{H}_I$. This is not surprising, however, if one recalls that, when all vertices are labeled identically $Q'=0$, i.e.

$$\sum_{i\neq j}\left(A_{ij}-\frac{d_i d_j}{2m}\right)=-\sum_{k=1}^{p}\frac{d_k^2}{2m}. \qquad (16)$$

These diagonal $\left\{d_k^2/2m\right\}$ terms provide a baseline weighting for terms in the external field, as described below.

Finally, the external field becomes

$$\mathcal{H}_E(x)=-2\sum_{k}\frac{d_k^2}{2m}\chi\left(x_k\right), \qquad (17)$$

where the minus sign makes the external field a penalization term. Therefore, the probabilistic model is given by

$$\mathbb{P}(x)\propto\exp\left\{\theta\left[-2\lambda\sum_{k}\frac{d_k^2}{2m}\chi\left(x_k\right)+\sum_{i\neq j}\left(A_{ij}-\frac{d_i d_j}{2m}\right)\left[2\delta\left(x_i,x_j\right)-1\right]\right]\right\}. \qquad (18)$$

The optimum $\arg\max\mathbb{P}(x)$ defines the optimal community structure for the network.

2.2 The Subnetwork-induced Probabilistic Model of Constrained Communities

In this section, we discuss an application of the constrained probabilistic model on a subnetwork (such as the giant connected component, for example) with the aim of perturbing a high-quality, unconstrained vertex labeling toward a high-quality, constraint-satisfying solution. We choose to present this here to elucidate our subsequent procedures for the optimization of Equation (18). We begin by discussing the subset-restricted probabilistic model.

Given a subset of vertices with indices S, the *vertex-induced* subgraph associated with S has adjacency matrix $A^{(s)}$ defined to be the sub-matrix of A induced by vertices with indices $k\in S$. The corresponding vertex degrees $d_k^{(s)}$ for $k=1,2,\ldots,|S|$ are computed as column totals of $A^{(s)}$. Finally, define

$$m^{(s)}=\frac{1}{2}\sum_{k\in S}d_k^{(s)} \qquad (19)$$

so that the S-restricted interaction component $\mathcal{H}_I^{(s)}(x)$ is defined as

$$\mathcal{H}_I^{(s)}(x) = \sum_{i \neq j \in S} \left(A_{ij}^{(s)} - \frac{d_i^{(s)} d_j^{(s)}}{2m^{(s)}} \right) \left[2\delta(x_i, x_j) - 1 \right]. \tag{20}$$

The S-restricted external field is given by

$$\mathcal{H}_E^{(s)}(x) = -2 \sum_{k \in S} \frac{\left(d_k^{(s)} \right)^2}{2m^{(s)}}. \tag{21}$$

As before in Equation (11), the optimal S-induced community structure is the mode of $\mathbb{P}^{(s)}(x)$, where

$$\mathbb{P}^{(s)}(x) \propto \exp\left\{ \theta \left[\lambda \mathcal{H}_E^{(s)}(x) + \mathcal{H}_I^{(s)}(x) \right] \right\}, \tag{22}$$

and is explicitly written as

$$\begin{aligned} \mathbb{P}^{(s)}(x) \quad \propto \quad \exp\Bigg\{ \theta \Bigg[& -2\lambda \sum_{k \in S} \frac{\left(d_k^{(s)} \right)^2}{2m^{(s)}} \chi(x_k) \\ & + \sum_{i \neq j \in S} \left(A_{ij}^{(s)} - \frac{d_i^{(s)} d_j^{(s)}}{2m^{(s)}} \right) \left[2\delta(x_i, x_j) - 1 \right] \Bigg] \Bigg\}. \end{aligned} \tag{23}$$

This is recognized as Equation (18) but with the superscript (s) to indicate the dependence on the set of indices S.

The ability to locally apply the constrained probabilistic model to a subset of vertices in a large network provides a powerful computation strategy. Suppose that x' is the vector of unconstrained community labels resulting from optimizing Equation (11) with $\lambda = 0$ over the entire network. One may view the constrained solution x'_c of community labels obtained by optimizing Equation (11) with $\lambda > 0$ as a perturbation of the unconstrained solution x'. Viewing it as such, one could consider locally perturbing x' directly as a computational device for approximating x'_c. We have found in applications, see Section 5, this often results in a constraint-satisfying, higher-modularity partition compared to a solution obtained via a direct application of the constrained model in Equation (11) with $\lambda > 0$ on large networks. This suggests the following three general strategies.

2.3 Three Strategies for Computing Constrained Optimal Communities

The three methods presented in the following are suggested procedures for obtaining a high-quality community structure through optimizing Equation (18).

2.3.1 The Direct Method

The *direct method* is the method of direct optimization of Equation (18) over the entire network. This may be achieved through a variety of discrete optimization procedures, however, we select a *Gibbs sampler* routine, discussed in Section 2.4, for this purpose. A Gibbs sampler routine is a Markov Chain Monte Carlo (MCMC) procedure for obtaining a sequence of samples from a joint distribution such as Equation (18). In particular, the samples produced are nearby modes of the probability distribution. At each step of a Gibbs sampler procedure, a single variable from the current configuration is re-drawn from the full conditional distribution of that variable given all other variables. In order to begin, a Gibbs sampler requires an initial configuration.

We initiate the Gibbs sampler in accordance with the Louvain method, see [16], in which each vertex is assigned a unique community label at initialization. In general, however, situations exist in which a different choice of initial condition is preferable. We shall encounter such a scenario in Section 2.3.2. Of course, the direct method may be applied with either $\lambda = 0$ or $\lambda > 0$ to achieve unconstrained or constrained communities, respectively.

Consider for a moment the probability function $\mathbb{P}(x)$ in Equation (18). Suppose that $\mathcal{L}$ is the set of possible community labels on a network of p vertices so that

$$\mathbb{P} : \mathcal{L}^p \mapsto [0, 1] . \tag{24}$$

The function $\mathbb{P}(x)$ may have a growing number of local maxima, i.e. modes, as the number of vertices p in the network increases. It is at these local maxima where the Gibbs sampler can find itself trapped and unlikely to recover. Thus, the results of several instances of the Gibbs sampler must be compared to alleviate this issue. That is, it is common for one to execute the optimization procedure multiple times in order to generate candidate optima and then to select the maximum from this set. This is potentially computationally prohibitive.

Therefore, the direct method with $\lambda > 0$ is our preferred method for constrained community detection on small networks so that the number of local maxima is manageable. In the case of a small network, (i) the number of local maxima is feasible and (ii) running several instances of a Gibbs sampler is tractable. We have found "small network" to mean when the number of vertices $|V|$ is, say, in the dozens. In this case, fewer and less expensive restarts of the Gibbs sampler are required to identify a high-quality community labeling.

Experimentally, it appears that the local maxima issue to be exaggerated by the inclusion of the external field in the formulation of Equation (18). Never-the-less, we have found that when $\lambda = 0$, so that the external field vanishes, the Gibbs sampler routine has improved performance. That is, identifying unconstrained communities with the direct method is more feasible and straight-forward. It is this insight which led us to explore our second proposed method for optimization of Equation (18) as discussed in the following.

2.3.2 The Local Perturbation Method

We have found direct application of the constrained model $(\lambda > 0)$ in Equation (18) on small networks to be effective. Moreover, we have found application of the unconstrained model $(\lambda = 0)$ on large networks to be feasible. Thus, we propose a simple strategy for perturbing the network-wide unconstrained solution into a constraint-satisfying solution. In short, the local perturbation method entails one application of the direct method on the entire network with $\lambda = 0$ and perhaps several applications of the direct method on subnetworks with $\lambda > 0$. Details on the local perturbation method are provided in the following.

Suppose that x' is the unconstrained vertex labeling as discovered by optimizing Equation (18) over the network with $\lambda = 0$. In other words, x' is the result of applying the direct method with $\lambda = 0$ to the entire network. Suppose that within x' there are $|C|$ unique labels, i.e. $|C|$ communities. We next identify the communities C_v in violation of the constraint $\mathcal{C}$ and if $C_v = \{\}$ then the constrained solution $x'_c = x'$, otherwise, some reconfiguration of x' is necessary to satisfy the constraint $\mathcal{C}$.

If reconfiguration of x' is necessary, then we iterate over each of the $|C_v|$ communities, applying the following procedure at each step k. To begin, suppose that c_k is a community in violation of the constraint $\mathcal{C}$. We identify a community c'_k with the property that if it were to relinquish a special vertex to community c_k, it would remain in compliance with $\mathcal{C}$, i.e. it currently contains at least two special vertices. Given c_k and having identified c'_k, we are prepared to execute *local reconfiguration*. Note: We have indicated a heuristic for selection of c'_k in Section 5.

We begin local reconfiguration by identifying the vertex indices S corresponding to all vertices contained in c_k or c'_k and apply the local method by optimizing Equation (22). The optimum x'_s of Equation (22) with $\lambda > 0$ results in a vertex labeling which locally satisfies the constraint. Recall that Equation (22) is simply a re-writing of Equation (18) and, therefore, we have ultimately just applied the direct method with $\lambda > 0$ to a subnetwork. We have found it favorable to initialize the Gibbs sampler with the vertex labels of $\vec{x}'$ corresponding to S. That is, when the Gibbs sampler is to begin operating on the subnetwork with vertex indices in S, it is initialized with a vertex labeling for the vertices with indices in S consistent with the unconstrained labeling computed previously for the entire network.

We overwrite the elements of x' with indices in S by the elements of x'_s to produce x^*. It should be noted that the change in Q' resulting from adjusting x' to x^* is

$$\Delta Q' = 2\sum_{i \neq j}\left(A_{ij} - \frac{d_i d_j}{2m}\right)\left[\delta\left(x'_i, x'_j\right) - \delta\left(x^*_i, x^*_j\right)\right] \quad (25)$$

$$= 2\sum_{i \neq j \in S}\left(A_{ij} - \frac{d_i d_j}{2m}\right)\left[\delta\left(x'_i, x'_j\right) - \delta\left(x^*_i, x^*_j\right)\right] \quad (26)$$

which indicates that the change in modularity for the entire network is a function solely of the terms in Q' for which both indices $i, j \in S$. At this point, the number of communities C_v in violation of the constraint $\mathcal{C}$ has been reduced by one. This procedure is executed for each $k = 1, 2, \ldots, |C_v|$ resulting in a vertex labeling x'_p which is a perturbation of the unconstrained solution x' and one which serves as a proxy for x'_c, the constrained solution to Equation (11).

To reiterate, the above method serves as a device for approximating the optimum x'_c of the constrained model, i.e. with $\lambda > 0$, by perturbing the optimum x' of the unconstrained model, i.e. when $\lambda = 0$.

2.3.3 The Hybrid Method

In the previous two sections we discussed the direct method which involves direct optimization of Equation (18) with $\lambda = 0$ for unconstrained communities and $\lambda > 0$ for constrained communities. To acquire constrained communities over the entire network, one can either (i) apply the direct method with $\lambda > 0$ at the outset or (ii) first apply the direct method with $\lambda = 0$ to derive unconstrained communities and then perturb the unconstrained communities using the local configuration method which itself involves application of the direct method with $\lambda > 0$ to a sequence of subnetworks. Again, we choose to perform optimization with a Gibbs sampler which requires initialization. In the direct method section, Section 2.3.1, we recommend initializing each vertex with its own label. Moreover, in the local perturbation method section, Section 2.3.2, when applying the direct method to a subnetwork, we recommend initializing each vertex to be in accordance with the unconstrained label previously acquired.

The *hybrid method* is simply a modification of the initialization of vertex labels during the local reconfiguration stage. Instead of initializing vertex labels as was done Section 2.3.2 during local reconfiguration, one could elect to initialize each vertex in the subnetwork with its own label, in effect discarding the information derived during the unconstrained community detection phase. The primary difference between the two competing initializations is that initializing each vertex in the subnetwork with its own label, as in the hybrid method, allows for more that two communities to be identified in the subnetwork. This is due to the fact that within the Gibbs sampler of Section 2.4, if a vertex has a label that is the last of its kind and it is reassigned to another label, the former label is out of existence and no other vertex will ever again be assigned it. Thus, the number of communities is a non-increasing function of the iteration step number of the Gibbs sampler. In total, the local perturbation method's initialization will discover at most two communities in the subnetwork, whereas, the hybrid method's initialization may discover more than two communities in the subnetwork. Never-the-less, it is unlikely that the two choices for initialization will produce different results.

2.3.4 Choice of Approach

In Sections 4 and 3 the networks considered have dozens of vertices and are said to be small. In this case, we utilize the direct method and directly compute constrained communities by optimizing Equation (11) with $\lambda > 0$. On the other hand, the hospital network of Section 5 has thousands of vertices, is considered large, and is susceptible to local optima traps. Thus, we employ the local perturbation procedure outlined above in which the unconstrained solution is obtained and then perturbed. The boundary between "small" and "large" networks depends on factors including computing power and network topology. As discussed in the following section, a discrete optimization such as modularity optimization is a hard problem. It is typical for one to employ several instances of a chosen computational optimization method to arrive at a set of proposed optimal solutions and then select the maximum thereof. Clearly, this is computationally prohibitive but is necessary in searching for a global maximum if the network topology dictates that the surface of Equation (18) has many local maxima. With a smaller network size, we have found that these two issues are not as restrictive as they are in larger networks and demonstrate our method in both scenarios in the following sections.

2.4 The Constrained Community Detection Estimation Algorithm and Gibbs Sampler Routine

From the expression $\mathbb{P}(x)$ in Equation (18) and the computation in Equation (14), it is straightforward to show that

$$\mathbb{P}(x_k|x_{-k}) \propto \exp\left\{2\theta\left[-\lambda\frac{d_k^2}{2m}\chi(x_k) + \sum_{j\neq k}^{p}\left(A_{kj} - \frac{d_k d_j}{2m}\right)\left[2\delta(x_k, x_j) - 1\right]\right]\right\}, \tag{27}$$

where $x_{-k} = (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_p)$ for $k = 1, 2, \ldots, p$. This closed-form expression for the full-conditional distribution of x_k is instrumental to our sampling/optimization procedure. Suppose that $\mathcal{L}$ is the set of all possible community labels for vertex v_k and define $\vec{p}_k$ to be the vector of probabilities computed from Equation (27) for each $x_k \in \mathcal{L}$. Evidently, the external field down-weights a label's probability of being assigned to vertex v_k if it is not consistent with the constraint. Finally, it is clear that $x_k|x_{-k} \sim \text{Multinomial}(\vec{p}_k)$. In the following we outline a Gibbs sampler to produce samples around the mode of $\mathbb{P}(x)$ for use in a simulated annealing optimization routine.

The starting point of our algorithm for obtaining the optimal community labels is an initial assignment of each vertex to its own community, i.e. vertex v_k is assigned community label k. Let $\mathcal{L}_t$ be the community labeling at time t and $\mathcal{L}_t'$ be a list of the unique community labels at time t. We define two loops: an outer loop which iterates over $t = 1, 2, \ldots, T$

and an inner loop which interates over $k = 1, 2, \ldots, p$, where, again, p is the number of vertices in the network.

Within the inner loop corresponding to generic vertex v_k at time t, iterate over all possible assignments $\mathcal{L}'_t$ and evaluate Equation (27) for each tentative assignment. Here, one may select a constant inverse temperature θ, as we have in selecting $\theta = 1$ in our applications. Let $\vec{p}_k^t$ be the vector of probabilities arising from this inner-most loop. We choose to reassign vertex v_k to the label which results from a single draw from the Multinomial $\left(\vec{p}_k^t\right)$ distribution since, as discussed earlier, $x_k|x_{-k} \sim$ Multinomial $\left(\vec{p}_k^t\right)$. For large T, the result of the algorithm converges to samples from the joint probability model in Equation (11) and, in particular, from around its mode. It should be noted that upon termination of the outermost loop, and hence the procedure, the number of communities has been identified. Akin to the initial phase of the Louvain method, see [16], the number of communities is reduced from p (one per each vertex) to the size of $\mathcal{L}'_T$.

The role of the external field is more explicitly evident in the formulation of Equation (27) and after the explanation of the sampling procedure above. For example, consider again the inner-most loop as described above in which the multinomial probabilities are computed. It is clear here that the external field down-weights a tentative label reassignment which does not move the solution toward feasibility. Intuitively, the external field provides a repulsive force away from a labeling which does not incorporate the external information or satisfy the constraint.

Modularity optimization is well-known to be a hard problem and, in particular, the *decision problem* version of modularity maximization is in the class NP-complete, see [15]. Strictly speaking, the problem of modularity optimization is stated as:
Modularity Optimization Problem: Determine a labeling x such that $\mathcal{H}(x) \geq \mathcal{H}(x')$ for all labelings x'.

Whereas, the decision problem version of modularity optimization is stated as:
Modularity Optimization Decision Problem: Given $K \in \mathbb{R}$, does there exist a labeling x such that $\mathcal{H}(x) \geq K$?

Existing approximate optimization procedures include greedy search, simulated annealing, and spectral methods, see [12]. While these methods are relatively efficient, there is no guarantee that the discovered community structure is optimal in the sense of modularity. Never-the-less, the discovered community is used in lieu of an optimal solution.

The above Gibbs sampling routine, like the methods listed above, outputs a high-quality solution in an efficient manner.

3. Developed Scenario

Suppose that G_n is a *stochastic block model* (SBM) on $n = a_1 + a_2 + 2$ vertices, a_1 in one block (Group #1), a_2 in the other (Group #2), and 2 special vertices. The within and between group probability of connection is p and $q = 1 - p$, respectively. Moreover, suppose that the two special vertices are connected to every other vertex in G_n with probability 1, including to each other.

Theorem 1 *If $a_1 < a_2$ and $p = 1$ $(q = 0)$ then (i) the unconstrained algorithm labels both special vertices as Group #1 and (ii) the constrained algorithm labels the special vertices as one in Group #1 and the other in Group #2.*

Proof 1 *See Section 7.3 in the Appendix.*

Alternate to the theorem above, if $p < 1$ $(q > 0)$ then there is a positive probability that the unconstrained solution will satisfy the constraint. Regardless, the constrained solution satisfies the constraint for any p, see Section 7.3 and Figure 1.

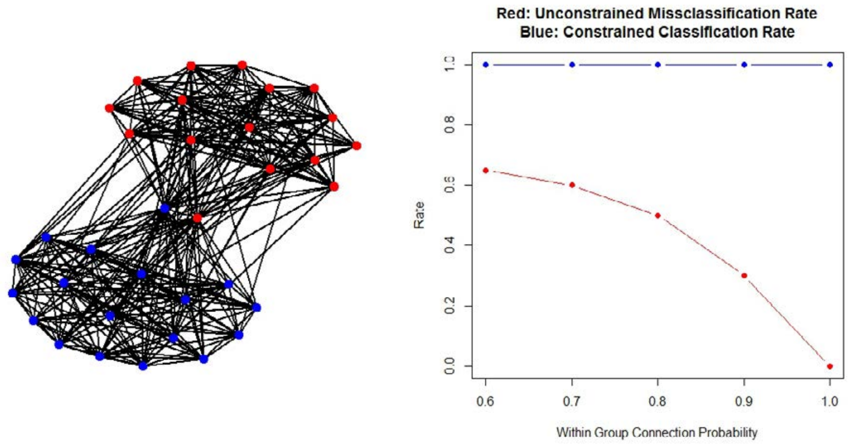

Figure 1: With $a_1 = 15$, $a_2 = 17$ and by setting $q = 1 - p$ and varying $p \in \{0.6, 0.7, 0.8, 0.9, 1.0\}$ we calculate the fraction of misclassified labelings of the special vertices by the unconstrained (red) and successful labeling of the special vertices by the constrained (blue) phases of our algorithm.

4. Illustrative Exercise

We present an exercises in the following which illustrate the role of the external field in our constrained community detection procedure. The illustration involves a network inferred from real, observational data. In this case, the network contains $|V| = 34$ vertices and is considered small. Therefore, we employ the direct method for community detection in the illustration, see Section 2.3.

Consider Zachary's observation of a karate club, see [19], in which a conflict arose between two members and the resulting fissure in the network was observed. Of the $p = 34$ participants in the club, the feuding members are represented in the network as vertices v_1 and v_{34} and, with the exception of a single vertex, Zachary's analysis correctly predicted community membership with a maximum flow - minimum cut algorithm. One could see here that our method for constrained community detection could be valuable in the sense that if we designate the vertices v_1 and v_{34} of the members in conflict as special vertices, then the external information dictates that one special vertex belongs in each discovered community.

Because the uninformed (unconstrained) modularity optimization procedure discovers communities in which vertices v_1 and v_{34} are labeled differently, satisfying the constraint, it is identical to the constrained optimum, see Figure 2.

To provide a more illustrative example with this same data, we investigate a scenario in which the karate club exhibits a further bifurcation. Recall that the members corresponding to vertices v_1 and v_{34} are feuding and the karate club split accordingly. Now, suppose that members corresponding to vertices v_1, v_{33}, and v_{34} are feuding (member number 33 has entered the dispute) and that we are interested in the allegiances stemming from the dispute given the current state of the network. In reflection of their dispute status, let vertices v_1, v_{33}, and v_{34} be special vertices so that each of the three discovered communities should contain at least one of them, see Figure 2. In this case, our constrained community detection algorithm discovered the optimal solution, see Figure 2, deviating from the unconstrained optimum.

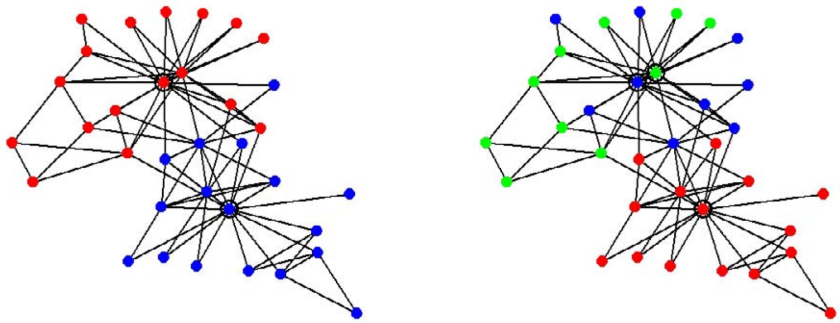

Figure 2: Left: Unconstrained community labels for the karate club network. Vertices v_1 and v_{34} are circled. Right: Constrained community detection after subsequent feud and split. Vertices v_1, v_{33}, and v_{34} are circled.

5. The Hospital Network

In this section, we employ our constrained community detection method on a network of hospitals to define health care communities (HCCs). We begin this section with background on a previous partition of the hospital network, i.e. health referral regions (HRRs), and conclude with a comparison between HRRs and our newly-defined HCCs.

5.1 Health Referral Region Background

As described in [17], hospital referral regions (HRRs) represent health care markets for tertiary medical care. Each HRR contains at least one hospital that performed major cardiovascular procedures and neurosurgery between 1992 and 1993. Three steps were taken to define HRRs and are summarized in the following. The first step was to identify the hospitals that performed at least ten major cardiovascular procedures in both years. The second step was to associate each of the remaining hospitals with a hospital from the first step according to where most of its patients received cardiac care. The third step was to reassign or merge groups primarily to achieve geographic continuity. These three steps resulted in a partition of the $3,436$ hospitals into 306 HRRs.

Assignment of hospitals to geographically proximal HRRs according to early-1990s data is logical since, according [18], the use of health care resources in the United States is highly localized. Now, nearly twenty-five years later, health care has modernized and, to an extent, become more globalized. To accommodate for this revolution in health care, we seek to group hospitals primarily according to the topology of the hospital network, as opposed to geographically. This is the primary discriminating attribute between the previously defined HRRs and our newly defined HCCs.

5.2 The Hospital Network Data

In an observational study of $p = 4734$ hospitals, the shared patient care activity for $n = 360,896$ Medicare participating physicians is obtained from insurance claims data. Each physician is assigned to the hospital where that physician provided inpatient services or where a plurality of that physicians patient panel had medical admissions, see [10]. Accordingly, the care activity for a hospital is defined to be the aggregation of the care activities of all physicians assigned to that hospital. More specifically, associated with each pair of physicians (“dyad”) involving the same patient(s) is a weight determined by the quantity of patient visits shared between the two physicians involving the same patient. We aggregate these weights to form a nationwide network of the $p = 4734$ hospitals, as described in the following.

5.3 Hospital Network Construction

Let B be the $p \times p$ matrix of aggregated shared visits. That is, the element B_{ij} is the total weighted care delivered to the same patients by physicians associated with hospital i and physicians associated with hospital j, see [18].

An edge between two hospitals is present in the network-graph representing the hospital network if the two hospitals share a sufficient level of shared visits. Specifically, let B_{ij} denote the total number of shared visits between hospitals i and j and define τ as

$$\tau = \text{Quantile}_{0.95}\left(B_{ij} : 1 \leq i < j \leq p\right), \tag{28}$$

see Figure 3. We define an adjacency matrix A according to

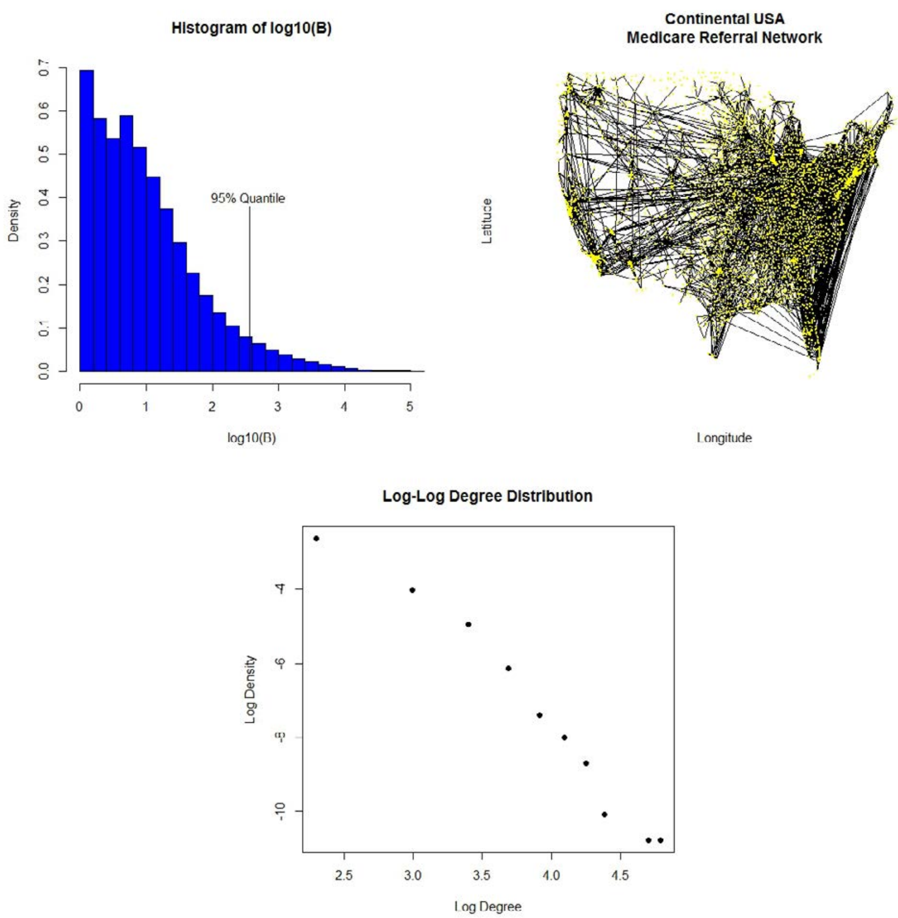

Figure 3: Top Left: Histogram of $\log\left(B_{ij}\right)$. Top Right: Network graph of the hospital referral network. Bottom: Log-log plot of the degree distribution of the hospital referral network.

$$A_{ij} = 1\left\{B_{ij} > \tau\right\} \tag{29}$$

so that the hospital network-graph is unweighted and undirected, see Figure 3. In addition to defining edges for those weights exceeding the threshold τ, we include edges corresponding to the maximum weight emanating from each vertex. The network-graph of the hospital network demonstrates a power-law degree distribution, see Figure 3, and has density $\delta = 0.0017$.

We define the constraint $\mathcal{C}$ to be that every health care community (HCC) contain at least one cardiac care facility, and equip the external field of the probabilistic model according to Section 2.1. Cardiac care facilities are those hospitals equipped for implantation of an implantable cardiac defibrillator (ICD) and are termed "ICD capable".

5.4 Constrained Community Detection of Health Care Communities

The full probabilistic model of Section 2.1 may be employed on this large network. However, as justified in Section 2.2, we recommend first computing unconstrained communities by setting $\lambda = 0$ in Equation (11) and then perturbing the solution by local reconfiguration on relevant subnetworks.

Due to the connectivity of the network, the discovered unconstrained $(\lambda = 0)$ communities tend to be geographically proximal, see Figure 4. In total, there exist $|C| = 98$ unconstrained communities with a distribution of sizes (number of vertices) and convex hull areas (square latitudes) as seen in Figure 4. The number of communities is determined through the identification of unconstrained communities by our procedure, see Section 2.4.

Of the $|C| = 98$ unconstrained communities, 17 do not contain a special vertex, a cardiac care facility. One such community is located in central Washington and is indicated in red in Figure 5.

At this stage we perturb the unconstrained solution to incorporate the constraint that each community contain at least one cardiac care facility. Naturally, if the target central Washington community in violation of the constraint is to gain a cardiac care facility, then it must be that another community loses one, assuming that it has one to lose, i.e. it started with at least two. In generating a list of candidate communities for relinquishing a cardiac care facility to the target community we note that, since switching a cardiac care facility from one community to another amounts to a sort of synthesis of the two communities, we should select the community (among those with at least two cardiac care facilities) that shares the weakest boundary with the target community. We define the weakness of a boundary as the local conductance, see [1], between communities S_i and S_j given by

$$cond_{ij} = \frac{\sum_{v \in S_i} \sum_{u \in S_j} A_{uv}}{a\left(S_i, S_j\right)}, \tag{30}$$

where A_{uv} is the $(u,v)^{th}$ element of the adjacency matrix for the network A and

$$a\left(S_i, S_j\right) = \min\left(\sum_{u \in S_i} \sum_{v \in S_i \cup S_j} A_{uv}, \sum_{u \in S_j} \sum_{v \in S_i \cup S_j} A_{uv}\right). \tag{31}$$

We utilize local conductance to assess the weakness of a boundary instead of local modularity due primarily to the fact that local modularity is not directly comparable across pairs of communities, see the Appendix. Moreover, conductance equals zero when the communities are perfectly separated and, therefore, provides a good metric for community boundary weakness.

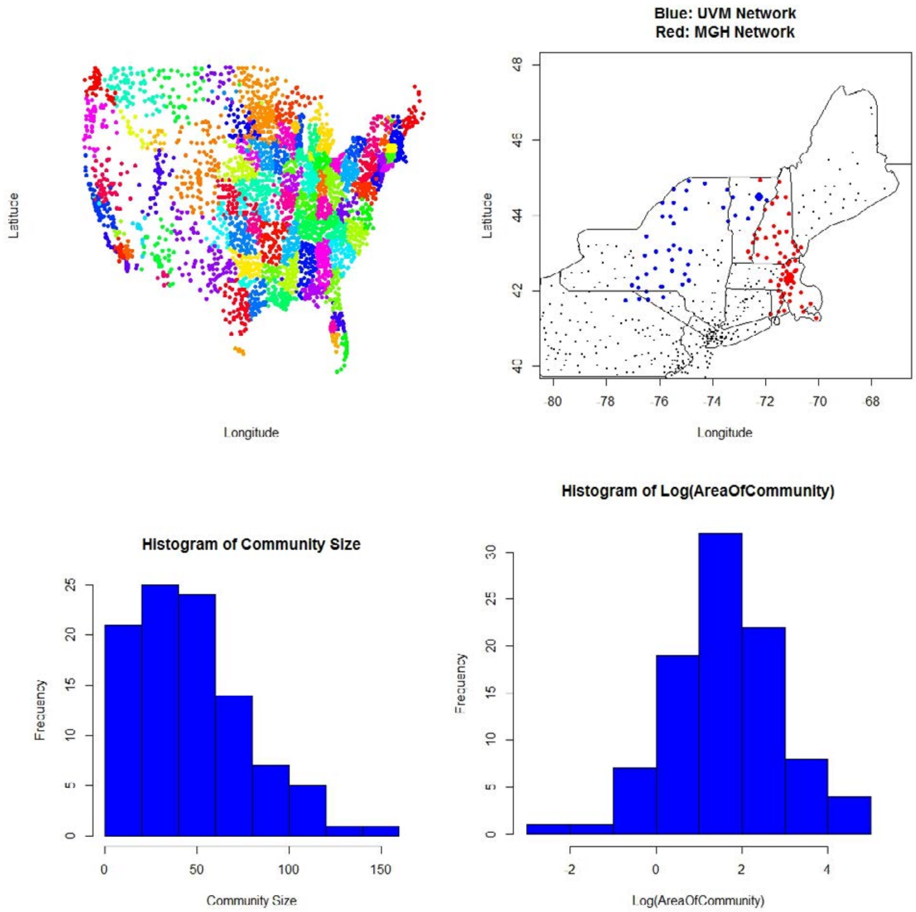

Figure 4: Top Left: Continental United States with hospitals colored according to community label. Top Right: Communities containing Massachusetts General Hospital (MGH) and University of Vermont Medical School (UVM). Bottom Left: Histogram of community sizes. Bottom Right: Histogram of log area of community, units of area in longitude by latitude

5.4.1 Example

Consider, for example, two candidates for relinquishing a cardiac care facility (i) the western Washington community, as indicated in blue in Figure 5, which shares edges with the central Washington community in violation of the constraint and (ii) a community in southern California which shares no edges with the community in violation. The central Washington community has a local conductance of $cond = 0.1$ with the western Washington community in contrast to a local conductance of $cond = 0$ with the southern California community. It turns out that $cond = 0.1$ is the maximum local conductance among all pairings of communities with central Washington and, therefore, the algorithm's next step is to merge the central Washington community with the western Washington community.

Now that the candidate community for integrating with central Washington has been established, we are prepared to execute our constrained community detection algorithm on the subgraph induced by the vertices of the two communities, see Section 2.2. We have coded our procedure to accept a "warm start" on the community labels. That is, unlike the unconstrained community detection procedure which initiated vertex labels arbitrarily, the vertices in the constrained community detection algorithm begins with the labels assigned

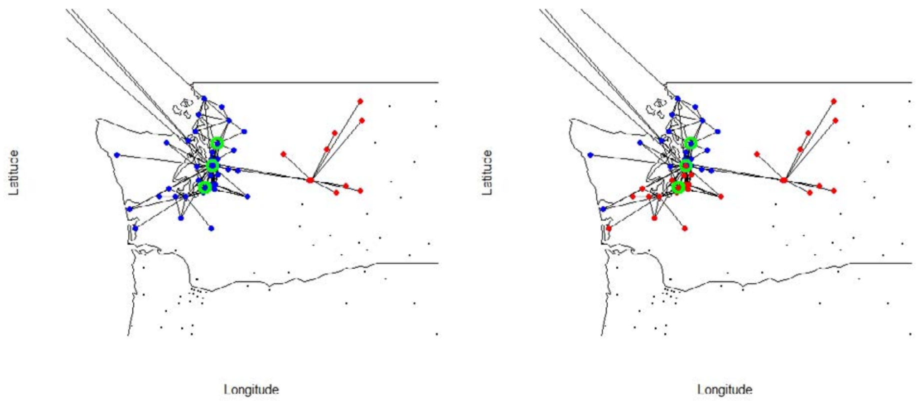

Figure 5: Left: Unconstrained communities. The red community does not contain an ICD capable hospital, indicated by green circles. Right: Constrained communities. The red community has incorporated western Washington and now contains an ICD capable hospital.

by the unconstrained phase. Now, with the penalty invoked, the labels are reassigned to the vertices in the subgraph by optimizing Equation (22). Upon termination, the unconstrained labels are overwritten by the newly discovered constrained labels. See Section 2.2 for a complete description of the above.

After having optimized the constrained model over the vertex-induced subgraph, one may note that the central Washington community acquired not one, but two special vertices. The constrained model enforces the minimal requirement that every community contain at least one special vertex. That two were acquired, however, is not entirely surprising. During the execution of the Gibbs sampler, see Section 2.4, a special vertex from western Washington community was relabeled from blue to red on account of the external field. Subsequently, the Gibbs sampler routine continues to draw samples from the full conditional probabilities in Equation (27) and, since the aforementioned special vertex had been reassigned to the central Washington community, several of its neighbors were reassigned, too, on account of the interaction effect. As a result, several members of the original western Washington community became members of the updated central Washington community, including the other special vertex.

Now that the central Washington community is in compliance with the constraint that it must contain at least one cardiac care facility, see Figure 5, we must repeat the above procedure for each of the remaining 16 communities remaining in violation of the constraint. It turns out that one of the remaining 16 communities is an isolated dyad and we do not impose the constraint in this case.

5.4.2 Change in Modularity

A key concern is how far the local reconfiguration method solution departs from the solution that would be obtained if the direct method could be feasibly applied. In either case, we begin by identifying the unconstrained ($\lambda = 0$) communities which are discovered solely on the basis of network connectivity. Recall that previously, when introducing the direct method, we stated that it may accept any initialization of vertex labels, including the unconstrained solution. We view the constrained solution as a perturbation of the unconstrained solution in which each community S' that does not contain a special vertex obtains one from another community S. Upon S' obtaining a special vertex from S, vertex label reassignment takes place to compensate for the switch. The difference between this and direct method on the entire network with the unconstrained solution as initialization, is that the direct method allows for all vertices in the network to participate in the label reassignment while the local reconfiguration permits only those vertices in S' and S to participate. Numerically, we have found the two competing procedures to produce constrained solutions (that are perturbations of the unconstrained solution) which are similar in terms of modularity score, but with the local perturbation method frequently out-performing its competing procedure. Moreover, suppose that S^* is a community not involved in the special vertex switching. Our experiments indicate that vertices in S^* remain unaltered in their labeling during the application of the constrained model to the full network. This provides further evidence for the utility of our local reconfiguration method for community label reassignment.

The remaining scenario to consider, following an application of the local reconfiguration procedure, arises when a community S' in violation does not neighbor a candidate community S containing a special vertex to dispense. In this case, one must directly recruit a special vertex from a community at a distance. Consider the change in modularity $\Delta Q'$ arising from relabeling a special vertex v_k from x_k to x'_k given by

$$\Delta Q' = 4 \sum_{j \neq k} \left(A_{jk} - \frac{d_j d_k}{2m} \right) \left[\delta\left(x_j, x'_k\right) - \delta\left(x_j, x_k\right) \right], \tag{32}$$

see Equation (14). As before, if S' is the community requiring a special vertex and S is the community to which a candidate special vertex belong, the above sum is further simplified to

$$\Delta Q' = 4 \left[\sum_{j \neq k : j \in S'} \left(A_{jk} - \frac{d_j d_k}{2m} \right) - \sum_{j \neq k : j \in S} \left(A_{jk} - \frac{d_j d_k}{2m} \right) \right]. \tag{33}$$

Hence, because the unconstrained solution is optimal in the sense of modularity, $\Delta Q' \leq 0$ and we select the special vertex which maximizes this quantity. In fact, since the community S' is fixed (community S depends on the special vertex v_k) during the selection of a special vertex, the first term $\Delta Q'_1$ in Equation (33) above may be ignored. The remaining term $\Delta Q'_2$ is approximately given by

$$\Delta Q'_2 \approx -4 d_k \left[1 - \frac{\sum_{j \neq k : j \in S} d_j}{2m} \right] \tag{34}$$

which provides intuition on how the special vertex is selected. Specifically, the special vertex to switch into S' which minimizes modularity loss will have (i) low degree and (ii) belong to a community S wherein vertices have high degree. Such a special vertex is an internal isolate within its own community. It was not necessary for us to select such internal isolates in defining HCCs. However, in general, one can easily foresee such a scenario arising.

Clearly the constrained solution obtains a lesser modularity value than the unconstrained solution. The extent to which imposing a constraint reduces modularity is depicted in Figure 6. The number of cardiac procedures for each ICD capable hospital is recorded and the minimum is 10. Certainly, with a large number of ICD capable hospitals, i.e. a large number of special vertices, the constrained solution is similar, if not identical, to the unconstrained solution and, thus, the resulting modularity of the constrained solution is approximately the modularity of the unconstrained solution. We set a lower threshold for the minimum number of cardiac procedures a hospital must perform in order to be considered an ICD capable hospital. As this lower threshold increases, the number of ICD capable hospitals decreases.

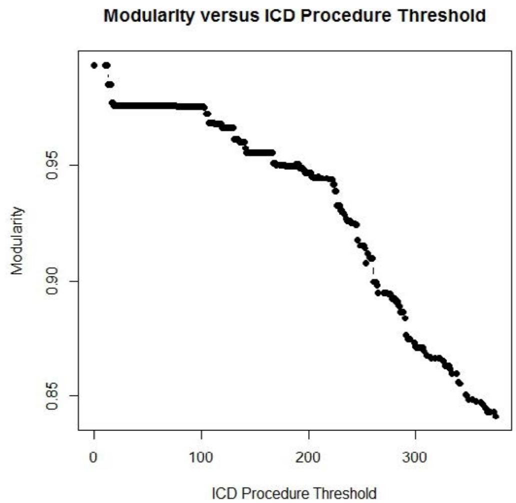

Figure 6: Decreasing modularity as a function of the cardiac procedures threshold. The threshold is the minimum number of cardiac procedures a hospital must perform to be classified as ICD capable. The y-axis represents the ratio of the modularity of the constrained solution to the modularity of the unconstrained solution. Note that modularity of unconstrained solution is an upper bound of the modularity of optimal constrained solutions.

5.5 HCC and HRR Comparison

An existing, non-network based partitioning of the network of hospitals is into health referral regions (HRRs), see Figure 7, of which there are 306, see [7] and [18].

The HRRs are approximately a refinement of the partition discovered by our method since $161/306 \approx 52.4\%$ of HRRs are wholly contained within an HCC, see Figure 7. We contend that our partition of hospitals into HCCs, being principally defined, represents a community structure more strongly coinciding with the topology, or homogeneity, of the referral network while, at the same time, preserving the constraint that every community contain at least one special vertex. This claim is substantiated by a contrast in modularity values $Q_{hrr} = 0.603$ and $Q_{hcc} = 0.812$, where Q_{hrr} is the modularity value resulting from the HRR partition and Q_{hcc} is the modularity value resulting from the HCC partition as discovered by our constrained community detection procedure.

Moreover, we consider the non-network-based measure of variation in guideline consistent usage of ICDs. Adherence to guidelines may be influenced by, for example, physician awareness, familiarity, and agreement with the guidelines, see [2].

Out of the 4734 hospitals included in our data, 1242 employ physicians performing ICD surgeries. For each of these hospitals where ICD surgeries took place, the fraction of guideline consistent use is determined. Our intention is to compare variation of guideline consistent ICD use between HCCs (and HRRs) and within HCCs (and HRRs). To that end, we compute the F statistics F_{hcc} and F_{hrr} for each of the two partitions where, in general, the F statistic is given by

$$F = \frac{\sum_i n_i \left(\bar{Y}_{i\cdot} - \bar{Y}\right)^2 / (K-1)}{\sum_{ij} \left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2 / (N-K)}, \tag{35}$$

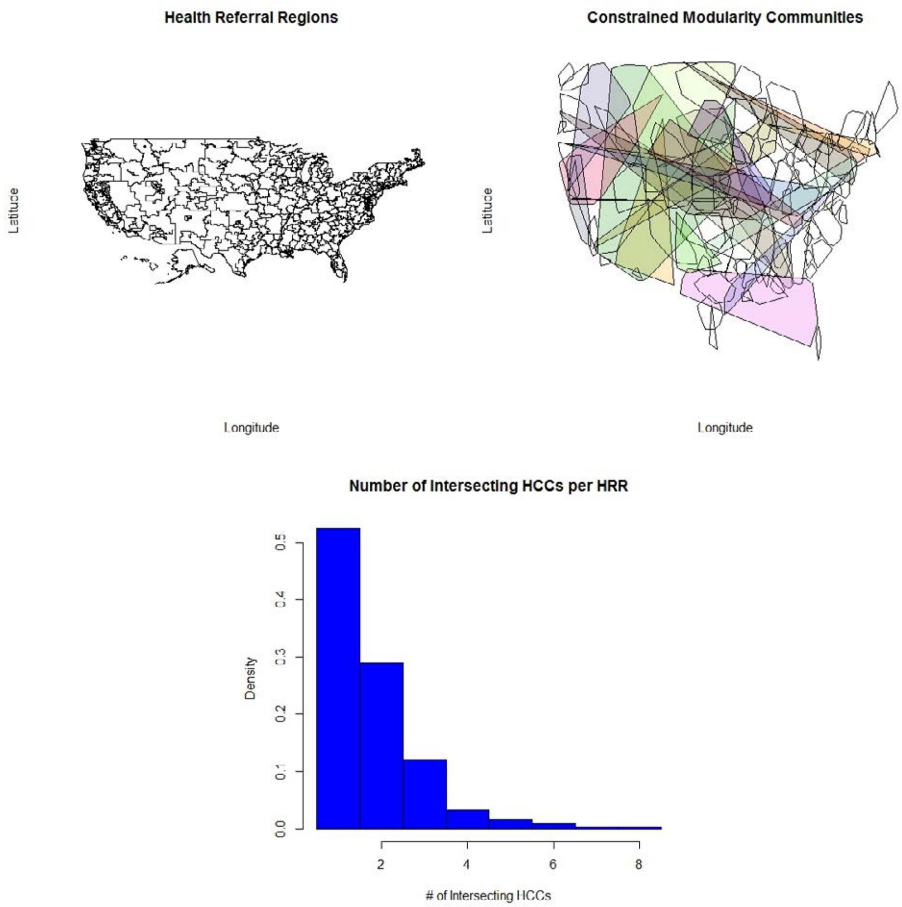

Figure 7: Top Left: Partition of the United States map into health referral regions. Top Right: Convex hulls of discovered health care communities. Because the communities are primarily based in network connectivity, there is no guarantee of geographic contiguity and thus some convex hulls span distances, as indicated by shading. Bottom: Histogram of HRRs as approximate subsets of HCCs.

where Y_{ij} for $i = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, n_i$ is the proportion of guideline consistent ICD use for hospital j in community i, $\bar{Y}_{i\cdot}$ is the mean of guideline consistent ICD usage in community i, $\bar{Y}$ is the mean guideline consistent ICD usage of all N hospitals, K is the number of communities, n_i is the number of hospitals in community i, and N is the number of hospitals with physicians who implant ICDs. We find that $F_{hcc} = 1.269$ $(p = 0.048)$ and $F_{hrr} = 0.828$ $(p = 0.8787)$. These results indicate that our community detection procedure has defined HCCs in such a way that the between community variability significantly exceeds within community variability despite the partition having no direct dependence on ICD use. However, the original definition of the HRRs does not permit such a claim.

6. Conclusion

Our method for informed discovery of optimally constrained communities is taken from the perspective of statistical physics and establishes the role of the external field in constraining the optimal solution to the space of feasible solutions.

There exists a disconnect between network science and health services research due in part to the incongruence between mathematical elegance and real-world constraints. The present article provides an illustration of the application of both a pure method and one with constraints. We solved the practical problem of partitioning a network of hospitals with the constraint that each community contain at least one specialized hospital, a cardiac care facility.

Though our method advances both the community detection literature and heath services literature, it is not complete from the perspective of a health care policy maker since many real-world constraints remain to be incorporated. For example, one may wish to, given the geographic locations of hospitals, define a specific diameter of the discovered communities. This paper is an initial foray into a line of thinking that we anticipate will substantially advance the practical utility of community detection.

In terms of health policy, further analysis of the discovered communities may lead to a better understanding of characteristics of hospitals within communities and detection of factors that drive variations in health care. This is the subject of our future work.

7. Appendix

7.1 Observed versus Expected Edges: Derivation of Modularity

In the course of comparing network characteristics, often the degree sequence is held fixed. If d_k is the degree of vertex v_k, i.e. the number of edges incident to vertex v_k, then $\mathbf{d} = (d_1, d_2, \dots, d_p)$ is the degree sequence of the network where p is the number of vertices in the network. The *configuration model* is a random graph model in which stubs, i.e. one end of an edge, from an existing network are randomly and independently re-wired to produce a new network with exactly the same degree sequence as the existing network. There are $2m$ stubs, i.e. $\sum_k d_k = 2m$, where m is the total number of edges in the network, so that the probability P_{ij} of vertices v_i and v_j to be connected in the configuration model is given by

$$P_{ij} = \frac{2d_i d_j}{2m\,(2m-1)}. \tag{36}$$

The expected number of edges between vertices v_i and v_j is therefore

$$E_{ij} = mP_{ij} = \frac{d_i d_j}{2m-1} \tag{37}$$

where m in mP_{ij} is present since there are m total re-wirings in the configuration model. However, for large m the quantity $2m-1 \approx 2m$ and, thus, we write

$$E_{ij} = \frac{d_i d_j}{2m}. \tag{38}$$

The approximation of $2m-1$ with $2m$ is favorable, too, in that it induces unbiasedness in the edge count. That is both

$$\sum_{ij} E_{ij} = \sum_{ij} \frac{d_i d_j}{2m} = 2m \tag{39}$$

and $\sum_{ij} A_{ij} = 2m$, where A_{ij} is the $(i,j)^{th}$ entry in the adjacency matrix for the network. Specifically, we consider an unweighted network in which $A_{ij} = 1$ if edge $\{v_i, v_j\}$ exists in the network and $A_{ij} = 0$, otherwise. In total, $A_{ij} - E_{ij}$ is precisely the number of observed edges minus the number of expected edges between vertices v_i and v_j.

Finally, suppose that we are interested specifically in partitioning the vertex set into two groups A and B. This gives rise to the modularity function

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \frac{s_i s_j + 1}{2}, \tag{40}$$

where $s_k = +1$ if vertex v_k belongs to group A and $s_k = -1$ if vertex v_k belongs to group B. Note that Q increases when $A_{ij} > d_i d_j / 2m$ and $v_i, v_j \in A$ or $v_i, v_j \in B$, i.e. when there exist more edges than expected between vertices v_i and v_j and when they are labeled similarly.

Since the vertex group labels are unknown, s_k for $k = 1, 2, \ldots, p$ are, too, unknown and are the arguments over which Q is maximized in order to determine the optimal grouping. It is straight-forward to show that, Q may alternately be written

$$Q = \frac{1}{4m} \sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2m} \right) s_i s_j, \tag{41}$$

Moreover, neither does the factor $1/4m$ and thus we define Q' as

$$Q' = \sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2m} \right) s_i s_j \tag{42}$$

which is recognizable as the Hamiltonian of an Ising model with interaction parameters $J_{ij} = A_{ij} - \frac{d_i d_j}{2m}$. More than two groups may be considered by redefining

$$Q' = \sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \left[2\delta\left(x_i, x_j\right) - 1 \right], \tag{43}$$

where x_k is the community label for vertex v_k and $\delta(\cdot, \cdot)$ is the Kronecker delta, at which point the Q' is recognized as the Hamiltonian of a Potts model.

7.2 Utility-based Construction of Modularity

We now consider an alternative determination of modularity. The authors of [13] define a utility function which explicitly penalizes/rewards a given partition, or community assignment, by considering the four possibilities of

$$\begin{array}{lcl} (a) & A_{ij}\delta\left(x_i, x_j\right) & \text{(within community edge),} \\ (b) & \left(1 - A_{ij}\right)\delta\left(x_i, x_j\right) & \text{(within community non-edge),} \\ (c) & A_{ij}\left(1 - \delta\left(x_i, x_j\right)\right) & \text{(between community edge), and} \\ (d) & \left(1 - A_{ij}\right)\left(1 - \delta\left(x_i, x_j\right)\right) & \text{(between community non-edge).} \end{array} \tag{44}$$

Clearly we favor (a) and (d) while opposing (b) and (c). The utility function is thus defined as

$$R = \sum_{ij} \left[a_{ij} A_{ij} \delta\left(x_i, x_j\right) - b_{ij} \left(1 - A_{ij}\right) \delta\left(x_i, x_j\right) \right. \tag{45}$$

$$\left. - c_{ij} A_{ij} \left(1 - \delta\left(x_i, x_j\right)\right) + d_{ij} \left(1 - A_{ij}\right)\left(1 - \delta\left(x_i, x_j\right)\right) \right], \tag{46}$$

for some parameterization of $(a_{ij}, b_{ij}, c_{ij}, d_{ij})$.

In the manner of [13], suppose that edges are weighted equally whether internal or external, i.e. $a_{ij} = c_{ij}$, and similarly for non-edges, i.e. $b_{ij} = d_{ij}$, so that

$$R = \sum_{ij} \left[a_{ij} A_{ij} - b_{ij}\left(1 - A_{ij}\right) \right] \left[2\delta\left(x_i, x_j\right) - 1 \right]. \tag{47}$$

A convenient choice for the remaining parameters is $a_{ij} = 1 - \gamma P_{ij}$ and $b_{ij} = \gamma P_{ij}$, for some $\gamma > 0$, so that

$$R = \sum_{ij} \left[A_{ij} - \gamma P_{ij} \right] \left[2\delta\left(x_i, x_j\right) - 1 \right]. \tag{48}$$

With this choice, in order for the total energy of edges to equal the total energy of non-edges, i.e. to satisfy

$$\sum_{ij} [A_{ij} - \gamma P_{ij}] = 0, \tag{49}$$

we have

$$\gamma = \frac{\sum_{ij} A_{ij}}{\sum_{ij} P_{ij}} = \frac{2m}{\sum_{ij} P_{ij}}, \tag{50}$$

so that

$$R = \sum_{ij} \left[A_{ij} - 2m \frac{P_{ij}}{\sum_{ij} P_{ij}} \right] [2\delta (x_i, x_j) - 1], \tag{51}$$

where it is recognized that $P_{ij} / \sum_{ij} P_{ij}$ forms a probability distribution.

If we simply choose $P_{ij} \propto d_i d_j$, i.e. edge probabilities depend only on vertex degree, then

$$R = \sum_{ij} \left[A_{ij} - \frac{d_i d_j}{2m} \right] [2\delta (x_i, x_j) - 1] = Q' \tag{52}$$

since $\sum_{ij} d_i d_j = 4m^2$. Clearly, we have arrived at the objective function Q' of Equation (43). Therefore, modularity is characterized (or most justified) when the true network model depends only on the degree distribution in the case of a single homogeneous group.

7.3 Illustration of Informational Capacity of Constraints

The network is comprised of $p = a + b$ standard vertices and 2 special vertices. The a vertices in group A and a vertices in group B are connected via a stochastic block model, see [11], with within connection probability p and between connection probability q. We choose to simulate the network with a stochastic block model for its explicit community structure.

The following is an extreme example. Suppose that $p = 1$ and $q = 0$ so that groups A and B are both cliques with no edges between them. The energy contribution of the interaction effect may be partitioned as

$$\begin{aligned} \mathcal{H}_I (s) &= \sum_{i \in A, j \in A} + \sum_{i \in B, j \in B} + 2 \sum_{i \in A, j \in B} \\ &\quad + \sum_{i \in A, j \in SV} + \sum_{i \in B, j \in SV} + \sum_{i \in SV, j \in SV}, \end{aligned} \tag{53}$$

where it is understood that $i \neq j$. It is clear from the first three sums, in the process of maximizing $\mathcal{H}_I$, that all vertices in A should be labeled A and similarly for B. The question remains, to which groups should the two special vertices be assigned?

Let S_4 be the fourth sum in Equation (53) and note that

$$S_4 = \sum_{i \in A, j \in SV} \left[1 - \frac{(a+1)(a+b+1)}{2m} \right] [2\delta (s_i, s_j) - 1]. \tag{54}$$

Let $S_4 (A, A)$ denote the value of S_4 when both special vertices are labeled A. Then

$$S_4 (A, A) = 2a \left[1 - \frac{(a+1)(a+b+1)}{2m} \right]. \tag{55}$$

On the other hand, if one special vertex is labeled A and the other labeled B then $S_4 (A, B) = 0$. Lastly, if both special vertices are labeled B then $S_4 (B, B) = -S_4 (A, A)$.

These observations together mean that the fourth and fifth sums in Equation (53), with community labels A for both special vertices, is computed as

$$\begin{aligned} S_4(A,A)+S_5(A,A) &= 2a\left[1-\frac{(a+1)(a+b+1)}{2m}\right] \\ &\quad -2b\left[1-\frac{(b+1)(a+b+1)}{2m}\right] \\ &= 2(a-b)+2\frac{(a+b+1)}{2m}\left[b(b+1)-a(a+1)\right]. \end{aligned} \tag{56}$$

Similarly, if both special vertices are labeled as B then

$$\begin{aligned} S_4(B,B)+S_5(B,B) &= -2a\left[1-\frac{(a+1)(a+b+1)}{2m}\right] \\ &\quad +2b\left[1-\frac{(b+1)(a+b+1)}{2m}\right] \\ &= -2(a-b)-2\frac{(a+b+1)}{2m}\left[b(b+1)-a(a+1)\right]. \end{aligned} \tag{57}$$

Note that

$$b(b+1)-a(a+1)=(b-a)(a+b+1) \tag{58}$$

so that

$$S_4(A,A)+S_5(A,A)=2(a-b)\left[1-\frac{(a+b+1)^2}{2m}\right] \tag{59}$$

and

$$S_4(B,B)+S_5(B,B)=-2(a-b)\left[1-\frac{(a+b+1)^2}{2m}\right]. \tag{60}$$

Note that

$$\begin{aligned} 2m &= a(a+1)+b(b+1)+2(a+b+1) \\ &= \left(a^2+b^2+2ab+2a+2b+1\right)-(2ab-a-b-1) \\ &= (a+b+1)^2-(2ab-a-b-1) \\ &\leq (a+b+1)^2 \end{aligned} \tag{61}$$

if $a,b>1$. It follows that

$$S_4(A,A)+S_5(A,A)\geq S_4(B,B)+S_5(B,B) \tag{62}$$

if $a<b$. Henceforth, it is understood that $a<b$.

Although, it did not enter into the above computations, the sixth sum in Equation (53) must now be considered and takes the values

$$S_6(A,A)=1-\frac{(a+b+1)^2}{2m} \tag{63}$$

and $S_6(A,B)=-S_6(A,A)$. This means that

$$\begin{aligned} \sum_{k=4}^{6}S_k(A,A)-S_6(A,B) &= 2(a-b)\left[1-\frac{(a+b+1)^2}{2m}\right] \\ &\quad +2\left[1-\frac{(a+b+1)^2}{2m}\right] \\ &= 2(a-b+1)\left[1-\frac{(a+b+1)^2}{2m}\right], \end{aligned} \tag{64}$$

which by previous results is ≥ 0 if $a - b + 1 \leq 0$ or, in other words, $a < b$. We have thus determined that the optimal, unconstrained community assignment for the special vertices when $p = 1$, $q = 0$, and $a < b$ is A.

In order to incorporate external knowledge that one special vertex should be labeled A and the other labeled B we must introduce into the computations the external field. Recall that the external field is given by

$$\mathcal{H}_E(s) = -2 \sum_k \frac{d_k^2}{2m} \chi(s_k). \tag{65}$$

Let $H_E(A, A) = -4\frac{(a+b+1)^2}{2m}$ and $H_E(A, B) = 0$ in accordance with the definitions of $\chi(s_k)$ and $\mathcal{H}_E$. Now, define

$$R(A, A) = \lambda H_E(A, A) + S_4(A, A) + S_5(A, A) + S_6(A, A) \tag{66}$$

and

$$R(A, B) = \lambda H_E(A, B) + S_4(A, B) + S_5(A, B) + S_6(A, B), \tag{67}$$

for comparison in the determination of λ for enforcing the constraint of our external information. Specifically, we want to know for what λ we have $R(A, B) \geq R(A, A)$.

We compute that

$$R(A, A) = -2\lambda \frac{(a+b+1)^2}{2m} + 2(a-b)\left[1 - \frac{(a+b+1)^2}{2m}\right] + \left[1 - \frac{(a+b+1)^2}{2m}\right] \tag{68}$$

and

$$R(A, B) = -\left[1 - \frac{(a+b+1)^2}{2m}\right] \tag{69}$$

which implies that

$$R(A, A) - R(A, B) = -2\lambda \frac{(a+b+1)^2}{2m} + 2(a-b+1)\left[1 - \frac{(a+b+1)^2}{2m}\right], \tag{70}$$

which is ≤ 0 precisely when

$$\lambda \geq \frac{(a-b+1)\left[1 - \frac{(a+b+1)^2}{2m}\right]}{\frac{(a+b+1)^2}{2m}}. \tag{71}$$

In the case that λ exceeds the lower bound above, the constraint is enforced and the two special vertices will be labeled differently.

7.4 Weakness of Boundary

In the course of enforcing our constraint in applying our constrained community detection procedure, we begin with a community in violation, i.e. it does not possess a cardiac care facility. We then identify all communities with at least two special vertices and rank them according to the weakness of their boundary with the community in violation. We rank according to the metric of local conductance and not local modularity for the reason illustrated below.

Consider a subnetwork consisting of two disjoint cliques of sizes a and b and that the community labels are such that all vertices in one clique are labeled A and all vertices in the other clique are labeled B. Note that this boundary is as strong as possible and should have a weakness of 0. We compute the (local) modularity of this subnetwork to be

$$\begin{aligned} Q &= \frac{1}{4m}\sum_{ij}\left(A_{ij}-\frac{d_i d_j}{2m}\right)\left[2\delta\left(x_i,x_j\right)-1\right] \\ &= \frac{1}{4m}\left[-a\frac{(a-1)^2}{2m}+a(a-1)\left(1-\frac{(a-1)^2}{2m}\right)+2ab\frac{(a-1)(b-1)}{2m}\right. \\ &\quad \left.-b\frac{(b-1)^2}{2m}+b(b-1)\left(1-\frac{(b-1)^2}{2m}\right)\right], \qquad (72) \end{aligned}$$

where $2m = a(a-1) + b(b-1)$. In the particular case that $a = b$ we have

$$\begin{aligned} Q &= \frac{1}{4m}\left[2a(a-1)\left(1-\frac{(a-1)^2}{2m}\right)+2a^2\frac{(a-1)^2}{2m}-2a\frac{(a-1)^2}{2m}\right] \qquad (73) \\ &= \frac{a(a-1)}{2m}, \qquad (74) \end{aligned}$$

which gives $Q = 1/2$ which is the maximum value over all choices of (a, b), see Figure 8.

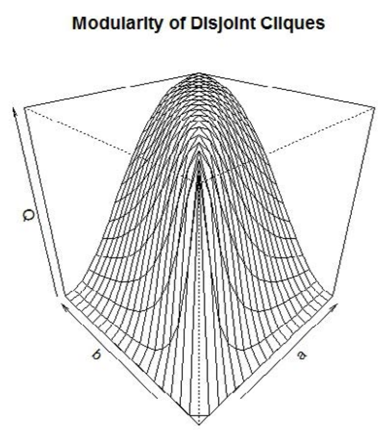

Figure 8: Modularity of the network in the disjoint cliques example for cliques of size a and b.

References

[1] Béla Bollobás. *Modern Graph Theory*. Springer-Verlag, 1998.

[2] Powe N.R. Wu A.W. Wilson M.H. Abboud P.C. Rubin H.R. Cabana M.D., Rand C.S. Why dont physicians follow clinical practice guidelines? a framework for improvement. *Journal of the American Medical Association*, 282(15):1458–1465, 1999.

[3] Fan Chung. *Spectral Graph Theory*. AMS and CBMS, Philadelphia, 1994.

[4] Eric Eaton and Rachael Mansbach. A spin-glass model for semi-supervised community detection. *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pages 900–906, 2012.

[5] I. Farkas G. Palla, I. Dernyi and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435:814–818, 2005.

[6] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821–7826, 2002.

[7] The Dartmouth Atlas Working Group. The dartmouth atlas of health care, June 2016. http://www.dartmouthatlas.org/data/region/.

[8] Bruce Hendrickson and Tamara G. Kolda. Graph partitioning models for parallel computing. *Parallel Computing*, 26:1519–1534, 2000.

[9] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift fr Physik*, 31:253–258, 1925.

[10] D. Gottlieb J. P. Bynum, E. Bernal-Delgado and E. Fisher. Assigning ambulatory patients and their physicians to hospitals: a method for obtaining population-based provider performance measurements. 42:45–62.

[11] Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83:016107, Jan 2011.

[12] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103:8577–8696, 2006.

[13] J. Reichardt and S. Bornholdt. Statistical mechanics of community detection. *Physical Review E*, 74(1), 2006.

[14] Leonardo B. L. Santos Charles N. de Santana Marcelo V. C. Diniz Thierry Petit Lobo Aristteles Gos-Neto Suani T. R. Pinho Roberto F. S. Andrade, Ivan C. Rocha-Neto and Charbel N. El-Hani. Detecting network communities: An application to phylogenetic analysis. *PLoS Comput Biol*, 7(5), 2011.

[15] Marco Gaertler Robert G”orke Martin Hoefer Zoran Nikoloski Ulrik Brandes, Daniel Delling and Dorothea Wagner. On modularity clustering. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 20(2):172–188, 2008.

[16] Renaud Lambiotte Etienne Lefebvre Vincent D Blondel, Jean-Loup Guillaume. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10, 2008.

[17] J. E. Wennberg. *Tracking Medicine: A Researcher's Quest to Understand Health Care*. OUP USA; 1 edition, 2010.

[18] J. E. Wennberg and M. M. Cooper. *The Dartmouth Atlas of health care*. American Health Association, Chicago, 1996.

[19] Wayne W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977.

[20] Yan Zhang, A. J. Friend, Amanda L. Traud, Mason A. Porter, James H. Fowler, and Peter J. Mucha. Community structure in congressional cosponsorship networks. *Physica A*, 387(7):1705–1712, 2007.